

A BESZÉD ARTIKULÁCIÓS SZINTÉZISE AZ ELTE FONETIKAI TANSZÉKÉN

Szűcs László
Eötvös Loránd Tudományegyetem
Fonetikai Tanszék

A beszéd fiziológiai folyamatainak megismerésére irányuló tudományos érdeklődés már a XVII. századtól kezdve összekapcsolódott az emberi beszéd(hangok) imitálására alkalmas szerkezet megalkotásának vágyával és előállítására tett kísérletekkel. Az emberi artikuláció során lezajló folyamatok (egyelőre még) mechanikus úton történő reprodukálása érdekében tett első igazán tudományos erőfeszítés a magyar származású Kempelen Farkas nevéhez fűződik. "Beszélőgépe", a korszak kiemelkedő jelentőségű, ám hosszú időig méltatlanul figyelmen kívül hagyott találmánya már képes volt a beszédhangokhoz hasonló hangzású akusztikai jeleket előállítani. Mechanikai megoldásokkal azonban elérhetetlennek bizonyult a biológiai hangképzés gyors és összetett folyamatának kielégítő imitálása. Az elektronika, a különböző rezgőkörök és hanggenerátorok alkalmazása forradalmasította a XX. század derekán az ez irányú próbálkozásokat, de az igazi áttörést a számítástechnikának a fonetikai vizsgálataiban, s ezzel párhuzamosan a beszédelőállító rendszerek létrehozásában való felhasználása jelentette.

Napjainkban, a magasfokú automatizálás korában, a tudomány és az ipar számos területén érezhetően megnőtt az igény egy igen precíz, a természetes intonációt is valósághűen visszaadó beszédsszimuláció alkalmazása iránt. Érthető hát, hogy a beszédelőállítás programja ma a nemzetközi és a hazai fonetikai kutatásokban is kiemelkedő szerepet kap.

A vizsgálatok jelenleg két fő megközelítési módszerre támaszkodnak, eszerint megkülönböztetünk akusztikus alapú és artikulációs alapú beszédsszintézist. A szintetizálás előkészítése magában foglalja a hangtest akusztikai szerkezetének, illetve a hang képzési (artikulációs) sajátosságainak alapos elemzését (beszédanalízis) is.

A világ számos kutatóintézetében többnyire a beszédhangok a k u s z t i k u s adatai alapján történő szintézis (formánsszintézis) az elterjedt módszer, amely eljárás lényege, hogy az egyes hangszeletek (szegmentumok) formáns-

jellemzőiből (időtartam, frekvencia, sáv szélesség, intenzitás) nyert adatbázis révén állítják elő a hangokat, szintetizátor vagy számítógép (digitál analóg átalakítás) segítségével. Hazánkban az akusztikus alapú szintetizáló rendszerek megalkotásában az első jelentős eredmények a 80-as évek elején születtek (Bolla 1982, Kiss--Olaszy 1982).

Az a r t i k u l á c i ó s alapú beszédelőállítás a hangképző szervek elektronikus modellálásának segítségével törekszik a szintézis megvalósítására. A kiindulópont itt a szintetizálandó hang(kapcsolat) artikulációs mozzanatát tükröző fiziológiai helyzet, melynek elemzése révén olyan adatokhoz jutunk, amelyekből az ún. toldalékcsőmodell-elmélet alapján kiszámíthatóak az akusztikus forma előállításához szükséges paraméterek, és végül az adott artikulációs helyzetnek megfelelő hangzást kapjuk eredményül. A beszédképzésben meghatározó szerepet játszó hangcsatorna (garat- és szájüreg) modellálására többféle megoldás is létezik. Mi a W. Nowakowska által kidolgozott szimulációs modellt használjuk (Nowakowska 1983). A csőmodell-elmélet szerint a hangcsatorna lényegében egy, a hangforrástól (larynx) az ajaknyílásig terjedő, változó keresztmetszetű csőként fogható fel, mely különböző átmérőjű, henger formájú szeletekre bontható. Ez a csőmodell kiegészül még az állandó geometriai struktúrájának tekinthető nazális csatornával (orrüreggel). Az így felállított modellen képzett hengerszeletek hosszúság- és sugáradataiból, a szupraglottális üregek rezonancia-tulajdonságait figyelembe véve, számítás útján meghatározható az akusztikus vetület.

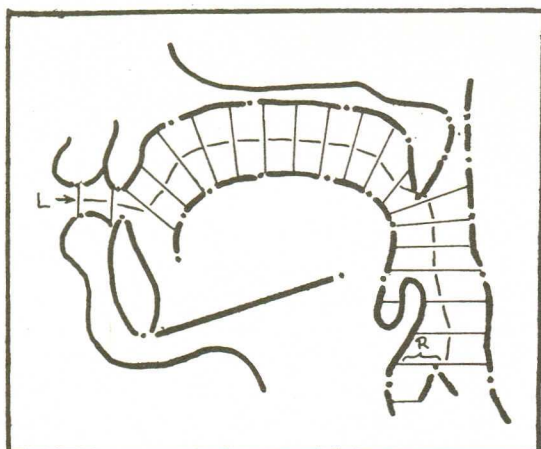
Az ELTE Fonetikai Tanszékének laboratóriumában, egy komplex fonetikai kutatás keretében, beszédanalízissel és -szintézissel egyaránt foglalkozunk. Bolla Kálmán eddig már 7 nyelv beszédképzése fiziológiai mechanizmusainak átfogó kinoröntgenográfiai elemzését végezte el (ld. MFF-hangalbumok). A beszédelőállítást illetően pedig mind az akusztikus, mind az artikulációs alapú szintézis körében folynak kísérletek. A továbbiakban ez utóbbiról kívánok részletesebben beszámolni.

Az artikulációs alapú hangszintézisre vonatkozó kutatások részét képezik annak az együttműködési munkakapcsolatnak, amely a 80-as évek elején jött létre az MTA Nyelvtudományi Intézete és az Institute of Fundamental Technological Research of the Polish Academy of Sciences között. A magyar fél munkálatai 1986 óta az ELTE Fonetikai Tanszékén folytatódnak, Bolla Kálmán és Földi Éva részvételével.

A csőmodell alapján kialakított paraméterek feldolgozásához szükséges számítógépes programot a varsói intézet munkatársai írták, a budapesti tanszék feladata az artikulációs helyzetet rögzítő kinoröntgenografikus felvételek elkészítése és adatolása, az adatbázis összeállítása, valamint az artikulációs sémák számítógépes ábrázolása, kirajzoltatása.

A vizsgálatok kiindulási alapját a Bolla Kálmán által készített több ezer kinoröntgeografikus felvétel képezi. Az elemzésre kerülő beszédhangokat e -- legkülönbözőbb artikulációs típusokat magában foglaló -- felvételsorozat alapján, természetes nyelvi környezetben (felsorolásban, mondatban, összefüggő szövegben) ejtett szavakból szegmentáljuk ki. Az adatbázis kialakítását a toldalékcsőben végbemennő artikulációs folyamatok vizsgálatára, valamint a fiziológiai mozzanatot rögzítő röntgenséma adatolásához már korábban kidolgozott számítógépes eljárások segítségével végezzük (ld. Bolla--Földi--Kincses MFF 15. 1986, 155--65.).

Az adatnyerés tehát a következőképpen történik. Az adott hang(kapcsolat) képzési folyamatáról készült felvettelt öt egyenlő időtartamú szeletre osztjuk. A szeletek kimerevített állóképeiről egy-egy, a toldalékcső kontúrjait tartalmazó (grafikai) sémát készítünk, majd az így kapott öt kép közül kiválasztjuk a hang tiszta fázisát leghűbben reprezentáló artikulációs konfigurációt. A kiválasztott rajzon a hangszalagoktól elindulva 1,5 cm-es távközönként metszeteket veszünk fel, a metszethatárokat a nyelvtől az ellentétes képzőszervre (garatfalra, kemény és lágy szájpadlásra) húzott merőleges egyenesek képezik. (A monitorként használt TV-készülék képernyőjén megjelenő artikulációs kép (röntgenogram) kb. egyharmadával adja vissza a valós méreteket, ezért nem 1, hanem 1,5 cm-es távközöket jelölünk ki.) Ezeket az egyeneseket megfelelve a toldalékcső adott metszeteire érvényes sugárértékeket kapjuk (mm-ben). Egy sémában a metszetek (és így a kapott adatok) száma, a toldalékcső hosszától függően kb. 17 és 24 között váltakozik. A metszethatárokon húzott merőlegesek és a toldalékcső kontúrjának (falának) metszéspontjait egy 320 x 200-as felbontású koordináta-rendszer segítségével, egy nagyon egyszerű és gyors módszerrel a számítógép ábrázolni tudjuk. A koordináta-adatokból a számítógép kiszámítja a sugárértékeket, az adatokat tárolja, képernyőre vagy nyomtatóra rajzolja a toldalékcső kontúrjait és a szeleteket tartalmazó sémát.



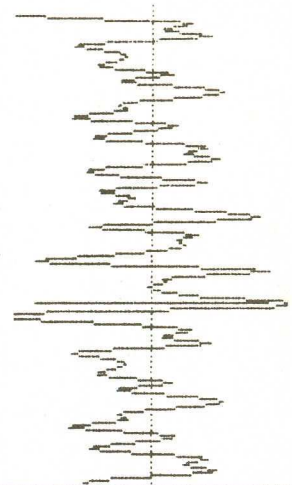
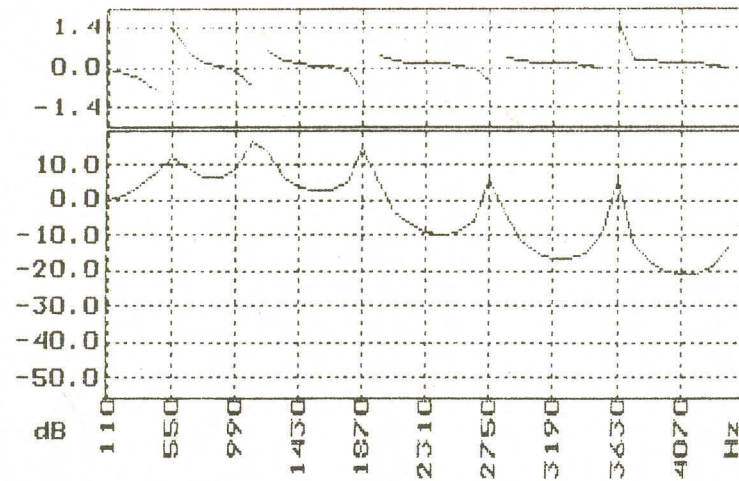
1. ábra A röntgenogram metszetei (L = a metszet hossza,
R = a metszet sugara cm-ben)

A lengyel kollégák által elkészített MOD nevű programnak a hang akusztikai vetületének kiszámításához a szegmentumok (szeletek) számára és az egyes szegmentumok sugáradataira van szükség. Az adatok átadása után a képernyőn (vagy printerrel kinyomtatva) megvizsgálható a szupraglottális üregek geometriai konfigurációja (alakzata), hang spektrogramja, továbbá leolvashatók a frekvencia- és intenzitásértékek, fázisviszonyok. A számítások után kapott akusztikai jellemzők (formánsszerkezet) alapján a program a hangot egy digitál analóg konverzió segítségével hallhatóvá teszi. Ha szükséges, a hangminőségen az adatok korrekciójával javíthatunk.

Az artikulációs alapú beszéd-szintézis megvalósítására készült számítógépes program az 50--5000 Hz közötti frekvenciaszerkezet ábrázolását teszi lehetővé (a vizsgálandó frekvenciasáv tetszés szerint változtatható), az intenzitásértékek 0--40 dB-es sávban olvashatóak le. Jelenlegi stádiumában a program az orális és a nazális vagy nazalizált magánhangzók szintézisét képes elvégezni.

Az eddig elért eredmények szerint lehetőségünk van minden magánhangzó-típus gépi megszólaltatására. Laboratóriumunk további feladata tehát a különböző mássalhangzók artikulációs alapú mesterséges előállításának megoldása.

NOS	USTA
	0.55
	1.10
	1.50
	1.35
	1.15
	1.05
	1.10
	1.15
	1.40
0.01	1.00
0.02	1.20
0.50	1.05
1.48	1.00
1.48	1.10
0.91	1.40
0.91	1.45
0.91	1.55
0.91	1.60
0.59	1.70
0.59	1.70
0.27	1.85



A magyar [ε] hang számítógépes diagramja

- BALÁZS János: Kempelen Farkas emlékezete. MFF 13. 1984, 11--9.
- BOLLA Kálmán: A magyar magánhangzók és rövid mássalhangzók képzési sajátosságainak dinamikus röntgenográfiai elemzése. MFF 8. 1981, 5--60.
- BOLLA Kálmán: Amagyar beszéd akusztikai szerkezetének analízise és szintézise. Kutatástörténeti áttekintés. MFF 10. 1982, 7--20.
- BOLLA Kálmán: Folyamatos beszéd szintetizáló rendszer magyar nyelven (VOXON). MFF 10. 1982, 118--27.
- BOLLA Kálmán--FÖLDI Éva--KINCSES Gyula: A toldalékcső artikulációs folyamatainak számítógépes vizsgálata. MFF 15. 1986, 155--65.
- KASSAI Ilona: Magyar beszéd fiziológiai vizsgálatok Kempelen óta. MFF 13. 1984, 39--42.
- KEMPELEN Farkas: Az emberi beszéd mechanizmusa, valamint a szerző beszélőgépeinek leírása. Ford. MOLLAY Károly. Budapest 1989.
- KISS Gábor--OLASZY Gábor: Interaktív beszéd szintetizáló rendszer számítógéppel és OVE III szintetizátorral. MFF 10. 1982, 21--45.
- MOLNÁR József: Kempelen Farkas beszélőgépe. MFF 13. 1984, 57--8.
- SUBOSITS István: Beszéd akusztika. Budapest, 1984, 13--7.